Osteoarthritis and Cartilage (2009) **17,** 281–283

doi:10.1016/j.joca.2008.09.008

Osteoarthritis and Cartilage

International Cartilage Repair Society

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Editorial
Blinding images to sequence in osteoarthritis: evidence from other diseases

D. T. Felson M.D., M.P.H.†‡*, M. C. Nevitt Ph.D.†‡

† *Clinical Epidemiology Research and Training Unit, Boston University, United States*

‡ *Department of Epidemiology and Biostatistics, University of California, San Francisco, United States*

Key words: X-ray, Image analysis, Osteoarthritis, Rheumatoid arthritis.

Osteoarthritis is a chronic disease that evolves slowly, and multiple longitudinal studies now ongoing are acquiring serial images of joints to track the structural progress of disease. When reading or analysing serial images using any method in which human judgment plays a part (as opposed to fully automated or computerized methods) it is generally accepted that viewing a subject's images from different timepoints grouped together reduces measurement error by enabling the readers to re-evaluate and adjust measurements made on these images, taking into account variation in film quality, positioning and other factors[1,2].

However, a major question that still arises is whether the grouped images from a person should be evaluated in known chronological order or whether the images ought to be presented with the analyst or reader blinded to the sequence in which they were acquired (the order of films is not revealed). (For the purposes of this editorial, we shall call the first approach 'known chronology' and the second 'blinded to sequence'.) The former scenario is similar to a clinical reading where a clinician or radiologist knows the sequence and tries to determine whether a patient has experienced a change in disease status. The latter procedure is one that has been instituted for some[3,4], but not all[5,6], clinical trials in osteoarthritis. Blinding to sequence is said to be conservative and rigorous and has been adopted by the osteoarthritis initiative in initial analyses of both radiographic and magnetic resonance images (MRI's).

The main rationale behind blinding to sequence is to reduce reader bias toward finding change in the expected direction and secondarily to minimize bias by readers who might be privy to information about risk factor status or treatment received. As noted above, these issues are relevant for any measurements from grouped serial images in which human judgment plays a role in determining the result, whether qualitative, semi-quantitative or quantitative. Even operations like segmentation of MRI images and digital measurements of joint space often have a manual component that can be influenced by whether the images are presented blinded to order or with known chronology. The potential for bias depends on which aspects of a method can be influenced by human judgment and to what extent.

The FDA has no formal policy on whether images from trials should be read blinded to sequence or with known chronology. In angiography studies as one example where serial images are used to measure the primary outcomes (luminal narrowing), images are analyzed in a manual or semi-automated fashion with known chronology. In a recent policy paper on best practices in Medical Imaging Techniques for trials[7], the FDA and PhRMA worked together to identify multiple different acceptable scenarios for image presentation including blinded to sequence or known chronology displays. As is noted, "the type of presentation often depends on the therapeutic area, the imaging technology being deployed, and the reasons for the review" of images. Given this uncertainty, we sought evidence on the performance of blinding to sequence vs reading with known chronology.

Evaluating the pros and cons of blinding to sequence is not easy. Ideally, readers or image analysts would have to take a set of images acquired sequentially in patients and analyze them both ways, one blinded to sequence and the other with known chronology in order to evaluate the effect of blinding to sequence on estimates of rates of change. Ideally, there also should be a gold standard definition to which the reading or image analysis can be compared, allowing an evaluation of tradeoffs between the effect of blinding on rates of change vs accuracy of outcome classification. For such a comparison to be of relevance to osteoarthritis, the images should have similar characteristics to either X-rays or MRIs of the knees or other joints. Specifically, they should be images in which planar measurements are made on a 2D image or series of slices with inferences made on where the edges of structures are in three dimensions and whether, over time, there has been a change in the position of these edges relative to one another. That would define a change in joint space or loss of cartilage.

We performed a MEDLINE search for articles published 1950 to current. Based on indexing of literature on blinding to sequence, we used indexing terms 'observer variation' or 'diagnostic error', these exploded terms were cross referenced with searches on the following content areas: spinal fractures, angiography, coronary angiography, rheumatoid arthritis (RA)/ra, duplex ultrasound carotid, osteoarthritis/ra (the suffix 'ra' refers to radiology studies). To find additional

*Address correspondence and reprint requests to: Dr D. T. Felson, M.D., M.P.H., Suite 200, 650 Albany Street, Boston University School of Medicine, Boston, MA 02118, United States. Tel: 1-617-638-5180; Fax: 1-617-638-5239; E-mail: ncarras@bu.edu

studies, we examined reference lists for the studies we found. After literature review, we found no studies evaluating reading blinded to sequence outside studies of spinal fractures and RA. To our knowledge, this has not been studied in osteoarthritis, and in vascular studies where serial images are read with known chronology, the issue has not arisen. Several studies, most of them of radiographs in RA, addressed these exact concerns.

The one non-RA study examined serial spine radiographs in patients with osteoporosis being followed for vertebral fractures. While this study included a reference reading to which the techniques were compared, it compared a morphometric measure of vertebral fractures from paired spine radiographs of known chronology to assessment of non-paired, randomly presented, single spine films and not films paired but blinded to sequence. Compared with reference assessment, false positive rates for vertebral fractures using single films were 20% vs only 7% when paired films were read with known chronology. While not directly examining the question of our focus, Ross and colleagues stated, 'We conclude that the assessments of X-rays for vertebral fractures in clinical trials should *not* be performed with the evaluator blinded to the sequence of the x rays[8]'.

Examples from studies of RA

Among studies evaluating progression of disease on serial radiographs in patients with RA, designs have varied. Most have compared the findings of paired readings either blinded or unblinded to sequence with respect to reliability and to how often progression is scored, but did not have a gold standard reading to which the different strategies could be compared.

Salaffi and Carotti[9] read films of the hands and wrists from 100 patients with RA at baseline and 18 months. Films were presented to each of two readers once with known chronology and at another reading session paired and blinded to sequence. The progression rates were higher when films were read with known chronology, but the readers showed slightly better agreement on progression when reading blinded to sequence.

In a study with a similar design[10] three readers each scored baseline and 12 month hand and foot radiographs of 284 patients who had participated in a randomized trial. When compared to readings that were paired but blinded to sequence, readings done with known chronology showed 30–50% higher progression rates and 5–10% higher standard deviations of those progression rates. The authors expressed concern about the higher standard deviation of progression rates and suggested that because the 'precision' of measurement was slightly greater with readings that were paired and blinded to sequence, these might be preferred. These authors did not take into account the higher progression rates seen when films were read with known chronology. It is not clear whether this higher standard deviation when films were read with known chronology meant greater variance of change or worse reader reliability.

In a similarly designed study evaluating hand and feet radiographs from 10 patients followed for a year and studied twice, van der Heijde and colleagues[2] had films scored by two experienced readers. They then conducted analyzes looking at whether there were differences in progression when examining films presented paired with known chronology vs blinded to sequence. They found a 50% higher rate of progression when films were read with known chronology. The standard deviation for progression was identical for readings that were blinded to sequence vs read with known chronology. The authors suggested that reading with known chronology yielded a higher signal to noise ratio for progression because of the higher rate of progression seen using this approach.

Thus, all of the studies in RA have shown higher rates of progression in patients where films were read with known chronology. The critical next question is whether the high progression rates seen in reading these images with known chronology are true cases of progression or represent false positives. If the former, then studies should be read with known chronology so as to increase the true signal detected in serial films. If the positives seen include many false positives, then blinding to sequence is needed to enforce a more conservative approach on readers. The only way of addressing this issue is to carry out a study with a 'gold standard' reading independent of the paired reading or to have an independent assessment of true vs false positive.

More recently, Bruynesteyn and colleagues conducted a study of hand and foot films from RA patients who had varying lengths of yearly follow ups. A group of experienced readers[11] reading the serial images with known chronology independent of the tested readers providing the gold standard measurements of whether progression was present. The films were scored using two different scoring systems widely used in RA, the Sharp/van der Heijde and Larsen methods. There were two readers for each method, and each of those readers read the paired films both blinded to sequence and, at a different time, with known chronology. Thus, in this study, there was both a clearcut gold standard reading and the ability to evaluate agreement between readers when they read serial films blinded to sequence or with known chronology.

In this study, interreader reliability was much higher when films were read with known chronology. As a result, the smallest detectable difference (SDD), was much larger for radiographs read blinded to sequence than reliability based on reading with known chronology. For example, when the films were read with known chronology, the SDD for the progression score was 5.0 and the mean progression score was 7.6, while for films read blinded to sequence, the SDD score was much higher (13.8), and the rate of progression was much lower in this group (4.5). Thus, if films were read blinded to sequence, few subjects would have had change beyond the SDD. Because of the poorer agreement of readers when films were read blinded to sequence and the lower rates of change seen when using this approach, sample size requirements for a dichotomous progression outcome would be much greater if films were read blinded to sequence than if the films were read with known chronology.

With respect to agreement with the gold standard group of readers, accuracy (based on the relationship between sensitivity and specificity) was slightly greater when films were read with known chronology. This was true even with the higher progression that was detected when films were so read. False positives were not created by this approach. The authors concluded that knowing the chronological sequence leads to an increase in detecting clinically relevant changes in patients without serious overestimation of non-relevant differences and that "analysing a clinical trial should be done preferably by reading films in chronological order."

The authors suggested further that reading images known to sequence would diminish sample size

Table I
Recommended approach to presenting serial images

Goal of study	Suggested approach to presenting images
Evaluating treatment efficacy	Known chronology*
Evaluating risk factors for disease	Known chronology
Evaluating and comparing methods for rates of incidence or progression	Blinded to sequence

*Exception if it is impossible to blind readers to treatment assignment or effects of treatment (e.g., trial with no control group).

requirements and make treatment effects easier to detect, issues of considerable concern in osteoarthritis where structural changes are modest over time.

Thus the evidence suggests that in RA the higher progression rates noted when films are presented with known chronology generally reflect true progression and that misclassification of progression does not occur at a higher rate when films are presented with known chronology. The higher signal compared with roughly equivalent noise translates into a better performance overall when films are presented with known chronology and that translates also into likely fewer subjects needed in studies and a greater likelihood of detecting treatment effects. While this may be true for radiographic reading in RA, it has not been studied in osteoarthritis and we can only speculate that it is of relevance.

The primary rationale behind blinding to sequence is to reduce reader bias toward finding change in the expected direction. Ironically, in a clinical trial in which outcome assessments are blinded to treatment assignment – a universally accepted standard – whatever bias there may be will be equal across treatment groups. If readers can be blinded to treatment assignment in a randomized trial, it is not necessary to blind them also to the sequence of the films (see Table I). If they are unaware of treatment assignment, their own biases cannot influence the trial results.

What then is the advantage of blinding to sequence and when should it be done? If there is evidence of the treatment on the image, then an outcome not based on this image may be preferable to using the image to assess outcome, as even presenting the image blinded to sequence is unlikely to negate this source of bias.

Further, if the question of interest is not about treatment efficacy in a trial or in the case of epidemiologic studies, about risk factors that increase the risk of progression but rather the focus is on *how often* progression or some other change occurs, then an argument can be made in favor of blinding to sequence (Table I). If films are read with known chronology, a reader could have the inclination to over-read progression (although in studies above that has not been found).

On the other hand, one could also argue that accuracy is more critical to determining rates of incidence and progression and if there is evidence that the risk of false positives is actually greater with blinding to sequence, then even studies focusing of rates of change could benefit from being done with known sequence. Finally, if the trial has no control group and one wants a conservative estimate of whether a treatment has affected imaging outcomes, then blinding to sequence may be preferable to reading with known chronology.

For just about every other question in which the rates of progression would be tied either to a treatment or a risk factor, blinding to sequence may compromises the ability of investigators to detect effects of interest. For risk factor studies, blinding to sequence may impair the investigators' ability to detect the effect of risk factors on disease incidence or progression. For treatments, blinding to sequence will make it more difficult to detect small treatment effects whose detection hinge on the 'greater accuracy' of unblinded assessments – sensitivity and power is increased because noise is reduced.

We conclude that, based on studies in RA, serial images from trials and longitudinal studies of osteoarthritis patients should usually be read with the images presented with known chronology. This is true for both semi-quantitative and quantitative evaluations that involve human judgments. The findings from these RA studies need to be confirmed in OA.

Conflict of interest

The authors are not aware of any conflicts of interest pertaining to this manuscript.

Acknowledgment

Supported by NIH AR47785.

References

1. Fries JF, Bloch DA, Sharp JT, McShane DJ, Spitz P, Bluhm GB, *et al*. Assessment of radiologic progression in rheumatoid arthritis. A randomized, controlled trial. Arthritis Rheum 1986;29(1):1–9.
2. van der HD, Boonen A, Boers M, Kostense P, van Der LS. Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials. Rheumatology 1999;38(12):1213–20.
3. Pavelka K, Gatterova J, Olejarova M, Machacek S, Giacovelli G, Rovati LC. Glucosamine sulfate use and delay of progression of knee osteoarthritis: a 3-year, randomized, placebo-controlled, double-blind study. Arch Intern Med 2002;162(18):2113–23.
4. Dougados M, Behier JM, Jolchine I, Calin A, van der HD, Olivieri I, *et al*. Efficacy of celecoxib, a cyclooxygenase 2-specific inhibitor, in the treatment of ankylosing spondylitis: a six-week controlled study with comparison against placebo and against a conventional nonsteroidal antiinflammatory drug. Arthritis Rheum 2001;44(1):180–5.
5. Brandt KD, Mazzuca SA. The randomized clinical trial of doxycycline in knee osteoarthritis: comment on the editorial by Dieppe. Arthritis Rheum 2006;54(2):684.
6. Bingham CO III, Buckland-Wright JC, Garnero P, Cohen SB, Dougados M, Adami S, *et al*. Risedronate decreases biochemical markers of cartilage degradation but does not decrease symptoms or slow radiographic progression in patients with medial compartment osteoarthritis of the knee: results of the two-year multinational knee osteoarthritis structural arthritis study. Arthritis Rheum 2006;54(11):3494–507.
7. Ford RMD. Report of Task Force II: best practices in the use of medical imaging techniques in clinical trials: consensus from a public meeting, October 16–17, 2007. Drug Info J 2008;42:515–23 (Ref Type: Generic).
8. Ross PD, Huang C, Karpf D, Lydick E, Coel M, Hirsch L, *et al*. Blinded reading of radiographs increases the frequency of errors in vertebral fracture detection. J Bone Miner Res 1996;11(11):1793–800.
9. Salaffi F, Carotti M. Interobserver variation in quantitative analysis of hand radiographs in rheumatoid arthritis: comparison of 3 different reading procedures. J Rheumatol 1997;24(10):2055–6.
10. Ferrara R, Priolo F, Cammisa M, Bacarini L, Cerase A, Pasero G, *et al*. Clinical trials in rheumatoid arthritis: methodological suggestions for assessing radiographs arising from the GRISAR Study. Gruppo Reumatologi Italiani Studio Artrite ReumatoideAnn Rheum Dis 1997;56(10):608–12
11. Bruynesteyn K, van der HD, Boers M, Saudan A, Peloso P, Paulus H, *et al*. Detecting radiological changes in rheumatoid arthritis that are considered important by clinical experts: influence of reading with or without known sequence. J Rheumatol 2002;29(11):2306–12.